TESTE T PARA AMOSTRAS INDEPENDENTES

Objetivo:

Testar se existe diferença entre a média de dois grupos independentes.

Nossa pergunta real será: *será que meus grupos são provenientes de amostras diferentes?*

Vimos que, para o teste t para uma amostra, a estatística era definida como:

$$t = \frac{média\ amostral - média\ populacional}{erro\ padrão}$$

Mas agora possuímos **dois** grupos amostrais, portanto podemos desenvolver a equação acima em função de **duas** amostras:

$$t = \frac{(méd\ amostral_1 - méd\ amostral_2) - (méd\ populacional_1 - méd\ populacional_2)}{estimativa\ do\ erro\ padrão}$$

A hipótese nula deve ser representada como hipótese de igualdade. Se nossa hipótese nula estiver correta, os dois grupos (amostras) foram retirados da mesma população, portanto:

H_0: média populacional$_1$ = média populacional$_2$

Sob a condição da hipótese nula, a equação acima se torna simplificada, pois se:

$média\ populacional_1 = média\ populacional_2$, logo,

$$média\ populacional_1 - média\ populacional_2 = 0$$

Assim, a estatística se torna:

$$t = \frac{(média\ amostral_1 - média\ amostral_2)}{estimativa\ do\ erro\ padrão}$$

A pergunta que resta agora é: *como obtemos o erro padrão da distribuição amostral da média de diferentes amostras?*

Para responder a essa pergunta, recorremos a Lei da soma das variâncias, que nos diz que:

A variância da distribuição amostral é igual a soma das variâncias das duas populações das quais as amostras foram retiradas.

Em outras palavras, se o erro padrão amostral é igual ao desvio padrão corrigido pela raiz quadrada do N amostral:

$$erro\ padrão\ amostral = \frac{desvio\ padrão\ amostral}{\sqrt{N}}$$

E agora possuímos duas amostras:

$$EP1 = \frac{desvio\ padrão\ amostral_1}{\sqrt{N_1}}\ e\ EP2 = \frac{desvio\ padrão\ amostral_2}{\sqrt{N_2}}$$

Precisamos expressar esses erros em função da variância, para enfim, podermos estimar o erro combinado dos dois grupos. Como:

$$desvio\ padrão = \sqrt{variância}$$

Assumimos a operação inversa:

$$variância = \ desvio\ padrão^2$$

Logo:

$$Var1\ = \left(\frac{desvio\ padrão\ amostral_1}{\sqrt{N_1}}\right)^2\ e\ Var2 = \left(\frac{desvio\ padrão\ amostral_2}{\sqrt{N_2}}\right)^2$$

Simplificando, temos:

$$Var1\ = \frac{desvio\ padrão\ amostral_1{}^2}{N_1}\ e\ Var2\ = \frac{desvio\ padrão\ amostral_2{}^2}{N_2}$$

Pela lei da soma das variâncias, podemos estimar o erro padrão da distribuição amostral da média de diferentes amostras:

$$Var\ = Var1 + Var2 = \frac{desvio\ padrão\ amostral_1{}^2}{N_1} + \frac{desvio\ padrão\ amostral_2{}^2}{N_2}$$

$$erro\ padrão\ = \sqrt{\frac{desvio\ padrão\ amostral_1{}^2}{N_1} + \frac{desvio\ padrão\ amostral_2{}^2}{N_2}}$$

Assim, a equação final da estatística se torna:

$$t = \frac{(média\ amostral_1 - média\ amostral_2)}{\sqrt{\frac{desvio\ padrão\ amostral_1{}^2}{N_1} + \frac{desvio\ padrão\ amostral_2{}^2}{N_2}}}$$

Se tomarmos duas amostras de tamanho N_1 e N_2 de uma população normalmente distribuída, a variável aleatória terá uma distribuição t com (N_1 + N_2 – 2) graus de liberdade. Em outras palavras, as probabilidades para aquela variável aleatória são iguais às áreas sob a curva t com GL = N_1 + N_2 – 2. Consequentemente, quando a população sendo amostrada é normalmente distribuída, nós podemos realizar um teste de hipóteses com a hipótese nula H_0: média populacional$_1$ = média populacional$_2$, empregando a variável aleatória t como nossa estatística de teste e usando a tabela da distribuição **t** para obter o valor crítico, **p**.

Exercício 1

Um trabalhador, suspeitando que o preço dos estacionamentos na região em que trabalha é **acima** da média da região em que ele mora, resolveu avaliar se sua suspeita era verdadeira. Ele coletou informações do preço de meia diária (12 h) em cinco estacionamentos próximos a seu trabalho e em cinco estacionamentos próximos a sua residência. Os resultados estão apresentados na tabela abaixo. Com base nessas informações, determine se a suspeita do indivíduo é verdadeira ou não. Assuma um α = 2,5% e considere que a normalidade e a homogeneidade foram respeitadas.

Custo do período de 12 h (em R$)	
Próximo ao trabalho	Próximo à residência
35,00	25,00
40,00	18,00
50,00	30,00
45,00	20,00
45,00	25,00

Exercício 2

Um dos estudos em psicologia social e educacional mais citados é de Rosenthal e Jacobson (1966), sobre como a expectativa dos professores pode influenciar o desempenho acadêmico das crianças. Rosenthal e Jacobson foram a uma escola básica e receberam a permissão de aplicar um teste de inteligência em todas as crianças. O teste havia sido desenvolvido recentemente e se baseava em uma inteligência não verbal que nenhum dos estudantes ou professores teria provavelmente visto antes. Assim, os pesquisadores foram capazes de criar expectativas nos professores, principalmente ao informar-lhes (erroneamente e de propósito) que o novo teste era "um teste para o desabrochar da inteligência". Os pesquisadores selecionaram aleatoriamente alguns alunos e informaram aos professores que aqueles haviam obtido desempenhos excepcionais no teste. Claramente, aqueles estudantes não haviam sido identificados com base em seus desempenhos reais. Em outras palavras, eles conduziram um experimento para testar as expectativas dos professores.

Oito meses depois, retornaram e administraram o mesmo teste novamente em todas as salas de aula. Sem grandes surpresas, a maioria das crianças apresentou um melhor desempenho nos testes, visto que completaram quase um ano de formação acadêmica. A pergunta principal, é claro, era se as crianças que haviam sido previamente rotuladas como "potencialmente inteligentes" apresentariam aumentos de desempenho ainda maiores do que as demais crianças (que não haviam sido rotuladas). A tabela abaixo apresenta informações resumidas de parte dos dados desse trabalho. Com base nessas informações, determine se as crianças rotuladas apresentaram um desempenho melhor do que as crianças não rotuladas. Assuma um α = 2,5% e considere que a normalidade e a homogeneidade foram respeitadas.

A criança foi rotulada	N	Média do desempenho	Desvio Padrão
Não	19	12,00	16,39
Sim	11	27,36	12,57

Exercício 3

Sempre assumindo a homocedasticidade dos dados, calcule o valor de t de Student para os seguintes valores de médias amostrais, desvio padrões amostrais e números de observações nas amostras, estabeleçando se podemos rejeitar ou não rejeitar a hipótese nula, sendo que:

a) Sempre assumindo 5% de nível de significância ($\alpha = 0.05$);

b) A hipótese nula é de que há igualdade das médias amostrais;

c) A hipótese alternativa é de que as médias são diferentes (teste t-bicaudal) entre as amostras.

I. X: Média=4.31, Desvio Padrão =2, Número de observações =16

Y: Média=2, Desvio Padrão =2, Número de observações =16

Hipótese nula: ( ) rejeita ( ) aceita

Valor de t encontrado: ___________________

Valor de t crítico a 5%: _________________

II. X: Média=4.31, Desvio Padrão =2, Número de observações =26

Y: Média=2, Desvio Padrão =2, Número de observações =26

Hipótese nula ( ) rejeita ( ) aceita

Valor de t encontrado: ___________________

Valor de t crítico a 5%: _________________

III. X: Média=4.31, Desvio Padrão =2, Número de observações =36

Y: Média=2, Desvio Padrão =2, Número de observações =36

Hipótese nula ( ) rejeita ( ) aceita

Valor de t encontrado: ___________________

Valor de t crítico a 5%: _________________

IV. X: Média=4.31, Desvio Padrão =2, Número de observações =30

Y: Média=2, Desvio Padrão =2, Número de observações =22

Hipótese nula ( ) rejeita ( ) aceita

Valor de t encontrado: ___________________

Valor de t crítico a 5%: _________________

V. X: Média=1007, Desvio Padrão =90.7, Número de observações = 31

Y: Média=1222, Desvio Padrão =70, Número de observações = 31

Hipótese nula ( ) rejeita ( ) aceita

Valor de t encontrado: ___________________

Valor de t crítico a 5%: _________________

HETEROCEDASTICIDADE

Quando aprendemos o teste t de Student, estabelecemos que haja igualdade nas médias populacionais e realizamos a estatística do teste, que tem como fórmula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Onde:

$$S_{X_1X_2} = \sqrt{\frac{(n_1 - 1)S_{X_1}^2 + (n_2 - 1)S_{X_2}^2}{n_1 + n_2 - 2}}$$

E os graus de liberdade → gl= $n_1 + n_2 - 2$

Mas neste caso é assumido que as duas amostras possuem variâncias (S_{X1}^2 e S_{X2}^2) significativamente iguais, ou seja: homocedásticas. Mas como saber se as amostras são heterocedásticas ou homocedásticas? Realizar o teste de Fisher (F), que tem uma fórmula muito simples:

$F_c=S_a^2/S_b^2$ → "crítico"

$F_T = (\alpha, gl_a, gl_b)$ → tabelado

$gl_a= n_a-1$; $gl_b= n_b-1$

α: probabilidade de erro tipo I.

Se $F_c<F_T$ → As amostras são homocedásticas (não há diferença significativa entre as variâncias).

$S_a^2 = S_b^2$

Se $F_c>F_T$ → As amostras são heterocedásticas (HÁ diferença significativa entre as variâncias).

$S_a^2 \neq S_b^2$

O que muda nos nossos testes?

A fórmula do teste t será:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{s_{\overline{X}_1 - \overline{X}_2}}$$

$$s_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

E os graus de liberdade (d.f.):

$$\text{d.f.} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

TESTE T PAREADO (MEDIDAS REPETIDAS)

Objetivo:

Testar se existe diferença entre duas médias de um mesmo grupo, avaliadas em dois momentos distintos.

Sempre os mesmos sujeitos!

Nossa pergunta será: *existe efeito do tempo (ou da intervenção) sobre a variável dependente contínua?*

Vimos que, para o teste t para uma amostra, a estatística era definida como:

$$t = \frac{média\ amostral - média\ populacional}{erro\ padrão}$$

Mas agora possuímos **dois** grupos amostrais (antes e depois), portanto podemos desenvolver a equação acima em função de **duas** amostras, avaliando a diferença entre o momento 1 e o momento 2 (ou momento 2 e momento 1):

$$t = \frac{(média\ da\ diferença\ amostral) - (diferença\ entre\ as\ médias\ populacionais)}{estimativa\ do\ erro\ padrão}$$

A mesma amostra é avaliada em dois momentos diferentes, e, portanto, é retirada da mesma população. Assim, o componente da diferença entre as médias populacionais pode ser removido da equação, pois é igual a zero:

$$t = \frac{média\ da\ diferença\ amostral}{\frac{desvio\ padrão\ da\ diferença}{\sqrt{N}}}$$

A hipótese nula deve ser representada como hipótese de igualdade. Se nossa hipótese nula estiver correta, não haverá diferença entre os dois momentos avaliados, ou seja:

H_0: média da diferença = 0

Se tomarmos uma amostra de tamanho N de uma população normalmente distribuída, a variável aleatória terá uma distribuição t com N – 1 graus de liberdade. Em outras palavras, as probabilidades para aquela variável aleatória são iguais às áreas sob a curva t com GL = N – 1. Consequentemente, quando a população sendo amostrada é normalmente distribuída, nós podemos realizar um teste de hipóteses com a hipótese nula H_0: média da diferença = 0, empregando a variável aleatória t como nossa estatística de teste e usando a tabela da distribuição **t** para obter o valor crítico, **p**.

Exercício 1

Um pesquisador desejava avaliar se um novo medicamento seria eficaz no tratamento da hipertensão. Ele selecionou, então, 10 pacientes e, para cada doente, foram registrados os valores de pressão arterial sistólica ao longo de uma semana, antes e depois do tratamento. A média dos valores semanais está apresentada na tabela a seguir. Com base nessas informações, determine se o tratamento foi eficaz. Assuma um $\alpha = 2{,}5\%$ e considere que a normalidade foi respeitada.

	Valores médios de pressão arterial sistólica (mmHg)	
Paciente	**Antes**	**Depois**
1	184	171
2	151	145
3	177	166
4	158	164
5	193	180
6	164	165
7	143	146
8	179	171
9	157	148
10	170	155

Exercício 2

Dois laboratórios determinaram a quantidade de cloro das mesmas amostras de água retiradas da rede de abastecimento de uma cidade. Com base nessas informações, há evidências suficientes para afirmar que existem diferenças significativas entre as medições realizadas pelos dois laboratórios? Assuma um $\alpha = 1\%$ e considere que a normalidade foi respeitada.

	Laboratório	
Amostra	A	B
1	1,15	1,00
2	1,86	1,90
3	0,75	0,90
4	1,82	1,80
5	1,14	1,20
6	1,65	1,70
7	1,90	1,95

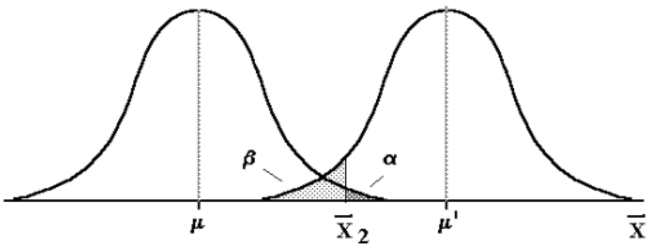

Tabela para teste t

α	0,200	0,150	0,100	0,050	0,025	0,020	0,015	0,010	0,005
*g.l.*1	1,37638	1,96261	3,07768	6,31375	12,70615	15,89447	21,20505	31,82096	63,65590
2	1,06066	1,38621	1,88562	2,91999	4,30266	4,84873	5,64280	6,96455	9,92499
3	0,97847	1,24978	1,63775	2,35336	3,18245	3,48191	3,89606	4,54071	5,84085
4	0,94096	1,18957	1,53321	2,13185	2,77645	2,99853	3,29763	3,74694	4,60408
5	0,91954	1,15577	1,47588	2,01505	2,57058	2,75651	3,00288	3,36493	4,03212
6	0,90570	1,13416	1,43976	1,94318	2,44691	2,61224	2,82893	3,14267	3,70743
7	0,89603	1,11916	1,41492	1,89458	2,36462	2,51675	2,71457	2,99795	3,49948
8	0,88889	1,10815	1,39682	1,85955	2,30601	2,44899	2,63381	2,89647	3,35538
9	0,88340	1,09972	1,38303	1,83311	2,26216	2,39844	2,57381	2,82143	3,24984
10	0,87906	1,09306	1,37218	1,81246	2,22814	2,35931	2,52749	2,76377	3,16926
11	0,87553	1,08767	1,36343	1,79588	2,20099	2,32814	2,49067	2,71808	3,10582
12	0,87261	1,08321	1,35622	1,78229	2,17881	2,30272	2,46070	2,68099	3,05454
13	0,87015	1,07947	1,35017	1,77093	2,16037	2,28160	2,43585	2,65030	3,01228
14	0,86805	1,07628	1,34503	1,76131	2,14479	2,26378	2,41490	2,62449	2,97685
15	0,86624	1,07353	1,34061	1,75305	2,13145	2,24854	2,39701	2,60248	2,94673
16	0,86467	1,07114	1,33676	1,74588	2,11990	2,23536	2,38155	2,58349	2,92079
17	0,86328	1,06903	1,33338	1,73961	2,10982	2,22384	2,36805	2,56694	2,89823
18	0,86205	1,06717	1,33039	1,73406	2,10092	2,21370	2,35618	2,55238	2,87844
19	0,86095	1,06551	1,32773	1,72913	2,09302	2,20470	2,34565	2,53948	2,86094
20	0,85996	1,06402	1,32534	1,72472	2,08596	2,19666	2,33625	2,52798	2,84534
21	0,85907	1,06267	1,32319	1,72074	2,07961	2,18943	2,32779	2,51765	2,83137
22	0,85827	1,06145	1,32124	1,71714	2,07388	2,18289	2,32016	2,50832	2,81876
23	0,85753	1,06034	1,31946	1,71387	2,06865	2,17696	2,31323	2,49987	2,80734
24	0,85686	1,05932	1,31784	1,71088	2,06390	2,17155	2,30692	2,49216	2,79695
25	0,85624	1,05838	1,31635	1,70814	2,05954	2,16659	2,30113	2,48510	2,78744
26	0,85567	1,05752	1,31497	1,70562	2,05553	2,16203	2,29581	2,47863	2,77872
27	0,85514	1,05673	1,31370	1,70329	2,05183	2,15782	2,29092	2,47266	2,77068
28	0,85465	1,05599	1,31253	1,70113	2,04841	2,15394	2,28638	2,46714	2,76326
29	0,85419	1,05530	1,31143	1,69913	2,04523	2,15033	2,28218	2,46202	2,75639
30	0,85377	1,05466	1,31042	1,69726	2,04227	2,14697	2,27827	2,45726	2,74998
35	0,85201	1,05202	1,30621	1,68957	2,03011	2,13316	2,26219	2,43772	2,72381
40	0,85070	1,05005	1,30308	1,68385	2,02107	2,12291	2,25027	2,42326	2,70446
45	0,84968	1,04852	1,30065	1,67943	2,01410	2,11500	2,24109	2,41212	2,68959
50	0,84887	1,04729	1,29871	1,67591	2,00856	2,10872	2,23378	2,40327	2,67779
60	0,84765	1,04547	1,29582	1,67065	2,00030	2,09936	2,22292	2,39012	2,66027
70	0,84679	1,04417	1,29376	1,66692	1,99444	2,09273	2,21523	2,38080	2,64790
80	0,84614	1,04319	1,29222	1,66413	1,99007	2,08778	2,20949	2,37387	2,63870
90	0,84563	1,04244	1,29103	1,66196	1,98667	2,08394	2,20504	2,36850	2,63157
100	0,84523	1,04184	1,29008	1,66023	1,98397	2,08088	2,20150	2,36421	2,62589
110	0,84490	1,04134	1,28930	1,65882	1,98177	2,07839	2,19860	2,36072	2,62127
120	0,84463	1,04093	1,28865	1,65765	1,97993	2,07631	2,19620	2,35783	2,61742
140	0,84420	1,04029	1,28763	1,65581	1,97706	2,07306	2,19244	2,35328	2,61140
160	0,84387	1,03980	1,28686	1,65443	1,97490	2,07063	2,18962	2,34988	2,60690
180	0,84362	1,03943	1,28627	1,65336	1,97323	2,06874	2,18743	2,34724	2,60341
200	0,84342	1,03913	1,28580	1,65251	1,97189	2,06723	2,18569	2,34513	2,60063
∞	0,84198	1,03697	1,28240	1,64638	1,96234	2,05643	2,17319	2,33008	2,58075